NOTES 06: TWO QUANTITATIVE VARIABLES

Stat 120 | Fall 2025

Prof Amanda Luby

When we're interested in the relationship between two quantitative variables, the best visualization is a **scatterplot**. If we want to summarize the relationship in a single number, we'll often choose the **correlation**.

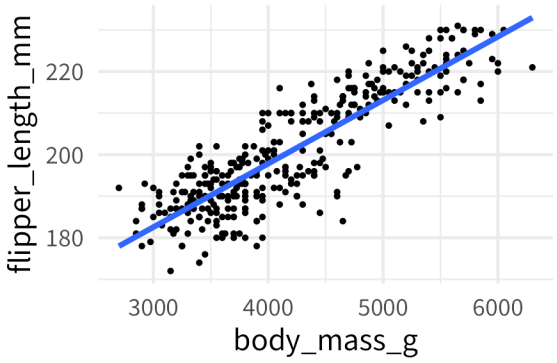
> **Note**
>
> Correlation

```
cor(penguins$body_mass_g, penguins$flipper_length_mm, use = "complete.obs")
```

```
[1] 0.8712
```

If the relationship is **linear**, we can also summarize the relationship with "the line of best fit" or "least squares" line.

```
ggplot(penguins, aes(x = body_mass_g, y = flipper_length_mm)) +
  geom_point(size = .5) +
  geom_smooth(method = "lm", se = FALSE)
```

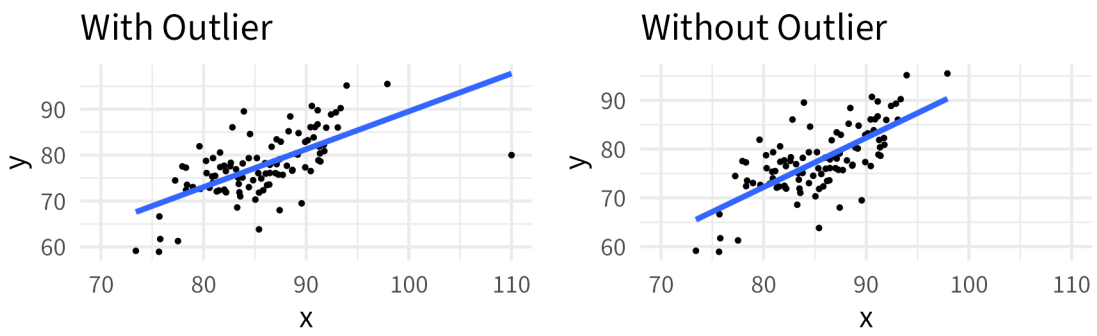



```
lm(flipper_length_mm ~ body_mass_g, data = penguins)
```

```
(Intercept) body_mass_g
  136.72956     0.01528
```

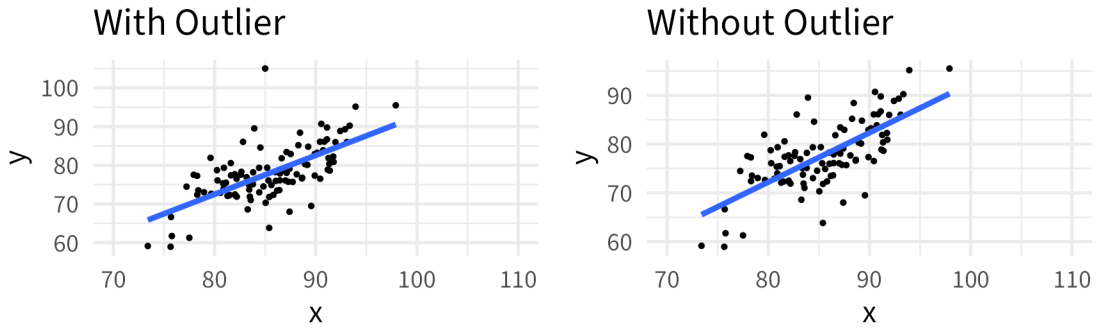
Interpretation:

Two types of outliers:



Best fit line with outlier: $\hat{y} = 7.1 + 0.82\,x$

Best fit line without outlier: $\hat{y} = -8.79 + 1.01\,x$



Best fit line with outlier: $\hat{y} = -8.1 + 1.01\,x$

Best fit line without outlier: $\hat{y} = -8.79 + 1.01\,x$